Scholar Publishing Group
International Journal of Multimedia Computing
https://doi.org/10.38007/IJMC.2021.020401
ISSN 2789-7168 Vol. 2, Issue 4: 1-8

E-commerce New Media Marketing Based on Data Technology

Xunyang Feng*

Jiangxi Vocational Technical College of Industry & Trade, Nanchang 330038, China

810911576@qq.com

**corresponding author*

Keywords: Data Mining, E-commerce, Online Marketing, Applied Research

Abstract: With the accelerating pace of life of Chinese residents, the expansion of the new media marketing market and the surge in the number of online shoppers have prompted e-commerce platforms to generate a wealth of e-commerce sales data. Faced with complex sales data, merchants need to analyze and understand the data in depth. Based on this, the purpose of this article is to study the application of e-commerce new media marketing based on data mining technology. This article first summarizes the basic theories of data mining, and then studies and analyzes its mining process, functions and methods. This paper systematically expounds the data preprocessing process, data preprocessing method selection and system algorithm realization of data mining technology in e-commerce new media marketing, and uses comparative method, observation method and other research forms to study the subject of this article. Experimental research shows that when user similarity calculation adopts the method Attribute-SimRank in this paper, the average hit rate of product recommendation for cluster center users is in most cases higher than that of SimRank and Attribute methods.

1. Introduction

In a society with explosive data growth, the number of companies that use e-commerce systems to use new media marketing for their businesses has soared [1-2]. At the same time, in the field of e-commerce, the application range of database management systems is also expanding. Enterprises have accumulated a large amount of data in their operations [3-4]. Reasonable analysis and prediction of these data can bring rich profits to business operators [5-6].

At present, in the development of informatization, many enterprises have many different enterprise application systems, and there are also great breakthroughs in the research of realizing enterprise data integration models [7-8]. Oracle Fusion Middleware is the sum of Oracle's series of application-based master's degree thesis 4 architecture products of Chengdu University of Technology [9-10]. It can realize the cooperation with suppliers who need middleware to complete seamless integration, supports hot swap, has stronger flexibility, and reduces implementation costs.

It uses the most cutting-edge software and hardware technology in the modern world to combine with the latest e-commerce technology to provide a simple and easy-to-run intelligent business management system [11-12].

This article aims at improving the efficiency of e-commerce marketing, and aims at the marketing of new e-commerce media based on data mining technology. By comparing the traditional e-commerce marketing system algorithm with the data mining technology-based e-commerce new media marketing algorithm designed and researched in this article, we can judge the feasibility of the research content of this article.

2. Research on E-commerce New Media Marketing Application Based on Data Mining Technology

2.1. Data Mining Process

(1) Determine the Business Object

Determine the mining purpose according to different needs, avoid wrong model selection and data source determination.

(2) Prepare Data

Data preparation occupies an important position in the mining process and directly affects the accuracy of the mining results.

(3) Data Mining

According to business requirements, mining purposes, and data characteristics, select appropriate algorithms, model parameters, and evaluation criteria, and extract model results from preprocessed data.

(4) Model Evaluation

Interpret and evaluate the results. Due to the problem of data selection or the selection of algorithm models, the patterns obtained by mining may be redundant or users are not interested. You need to go back to the first step and re-mining to obtain more valuable knowledge that meets the needs of users.

(5) Results Performance and Knowledge Assimilation

Use intuitive diagrams for mining modes to users, such as charts, graphs, etc.; and integrate the knowledge obtained from the analysis with the business information system.

2.2. Analysis of E-commerce Marketing System Based on Data Mining Technology

(1) System Architecture Design

This system adopts B/S structure and MVC structure hierarchical design. Consider the stability of the system. Security and scalability, using JavaEE technology and SSH framework to achieve. The system is divided into interface display layer, access control layer, business logic layer and data access layer.

(2) System Module Realization

Before the e-commerce marketing system, the business data should be transformed and integrated, and the analysis and prediction should be carried out based on the analysis and prediction. The system consists of a data processing module, a sales analysis module and a sales forecast module.

1) Analysis of data preprocessing module

Data collection

The main source of data is the sales data of each company. The information in the sample data includes: sales-related (orders, return orders, products, prices, sales dates, company numbers); basic attributes of products (product ID, category, color, size, unit price, discount); profit-related attributes (purchase price, sales price, profit); enterprise-related attributes (enterprise number, superior company, region); time-related attributes (year, season, month).

Data cleaning

Remove irrelevant information-customer information in the order, large receipt information, remarks, etc. are not related to the sales analysis, forecast and profit analysis of the product, and delete it; clean dirty data-for example, in the product classification, due to different operator input of "knitwear" and "sweater/knitwear" as two different categories, need to be processed, unified input value: In addition, for deviating from the conventional, scattered and a small number of tuples, delete based on experience; detect outliers by constraining the field range to determine the rationality of the data; vacant value processing-this article uses the sample average value to fill. For example: if the sales date is vacant, fill it in according to the average sales date of the season the product belongs to.

2) Analysis of sales forecasting module

The sales forecasting module is mainly for forecasting the company's product sales and business turnover. Commodity sales forecasting can enable operators to pre-estimate the purchase quantity and distribution quantity of goods, effectively solve the problem of frequent allocation among enterprises, and reduce inventory backlog problems. Turnover forecasting enables operators to understand the business trend of the enterprise in a macroscopic view.

This module is implemented using multiple linear regression analysis algorithm. Similar to cluster analysis, when performing regression analysis, data preparation is first required. Select each company's business data related to sales, and collect them into a relatively complete data table. Since the sales system concentrates on analyzing the sales of goods, the data of cluster analysis and regression analysis are all related to sales orders, product tables, return tables, enterprise tables, etc., so the regression analysis data table and the cluster data table use the same one. Data table to avoid repeated data preprocessing work.

3) Sales analysis module design

The main purpose of cluster analysis is to divide data into the same category according to profit and product sales. After clustering, the objects have similar sales proportions in similar profit margins, that is, they have more common points. The sales analysis module analyzes the company's sales and profits, and obtains the profit sales division, to provide a certain method for solving the problem of high sales and low profits. Using the improved algorithm, the initial mean value needs to be selected in advance and automatically divided into classes.

2.3. Customer Segmentation Based on Structural Similarity and Attribute Similarity

The similarity or distance between objects is one of the very important concepts in data mining tasks. In order to obtain valuable patterns and rules from massive data, it is the first problem to be solved accurately and effectively to measure the similarity or difference of data objects. Similarity calculation has been widely used in many fields such as information retrieval, citation analysis, recommendation system and community discovery.

The similarity calculation method proposed in this chapter is mainly applied to relational data mining, so the following will first briefly introduce the research status of relational data mining, and

then focus on the SimRank method and attribute feature similarity calculation method used in this chapter.

(1) Structural Similarity Calculation Method Simrank

Objects A and B in the same data space are more like objects A and B because they are simultaneously associated with object C or D in another data space, and the distance between them is closer. The same is true of the assumptions underlying the SimRank algorithm. The algorithm first represents the data object in the relational data as a node in the graph theory model, and the relationship between the objects is represented as the directed connecting edge between the nodes. The similarity between the nodes can be obtained using the SimRank calculation formula, as follows Formula (1) shows:

$$S(a,b)=\frac{C}{|I(a)||I(b)|}\sum_{i=1}^{|I(a)||I(b)|}\sum_{j=1}^{}S_{link}(I_i(a),I_j(b)),(a\neq b) \tag{1}$$

Formula (1) can calculate the structural similarity score between any node pair a and b after multiple iterations until convergence.

(2) Attribute Feature Distance Measurement

This paper divides all attributes into numerical attributes and categorical attributes when measuring the attribute feature distance, and uses different measurement methods for different types of attributes. For numerical attributes, in order to avoid the influence of data dimension, the attribute value is generally standardized, and the value range of the attribute is mapped to the same interval such as [0-1]. The formula for calculating the attribute similarity between objects is shown in formula (2).

$$Sim_A(x_{ij},x_{ik})=e^{-\|(x_{ij}\cdot^A-x_{ik}\cdot^A)\|} \tag{2}$$

In formula (2), $^{\|(x_{ij}\cdot^A-x_{ik}\cdot^A)\|}$ represents the attribute distance between objects, that is, the sum of the differences of the attributes of the objects.

(3) Evaluation Index of Clustering Result

In order to verify the validity of the similarity calculation method proposed in this chapter, we indirectly verify the authenticity and reliability of the Attribute-SimRank similarity calculation by objectively evaluating the clustering quality. In this article, the objective evaluation index of clustering we use is compactness, and the calculation of compactness is as follows

$$C_f=\frac{\sum_{i=1}^{K}\sum_{x\in C_i}d(x,a_i)}{\sum_{1\le i\le j\le K}d(a_i,a_j)} \tag{3}$$

$$d(a_i,a_j)=1-S_{ASimRank}(a_i,a_j) \tag{4}$$

Compactness is defined as the ratio of compactness within a cluster to isolation between clusters. A better clustering result is generally that the tightness within the cluster is shown by the small distance between objects within the cluster, and the greater isolation between clusters is shown by the greater distance between the centers of each cluster.

3. Experimental Research on E-commerce New Media Marketing Based on Data Mining

3.1. Experimental Protocol

In order to make the experiment more scientific and effective, this experiment uses

Attribute-SimRank similarity to improve the K-medodis clustering algorithm and conducts experiments on three data sets. The number of iterations of K-medodis is set to 1000. The input of K-medodis clustering is the similarity matrix and the number of clusters of the objects to be clustered, and the output is the cluster label of each object. The clustering results will be different when the number of clusters is set differently. Therefore, we have carried out grouping experiments on different numbers of clusters. Two sets of experiments were carried out when the Product data set was running. The number of clusters was set to 8 and 10, and the Movie data set the number of runtime clusters is 5 and 8, and the number of runtime clusters in the Customer data set varies from 5 to 10, and a total of 6 sets of experiments are carried out.

3.2. Research Methods

(1) Comparative Analysis Method

In this experiment, a total of six groups of experiments were set up for comparative analysis, and the data obtained were analyzed and counted. These data not only provide theoretical support for the topic selection of this article, but also provide data support for the final research results of this article.

(2) Observation Method

This experiment observes and records data on the hit rate of product recommendation among three similarity methods favorites of Kun Center users. These data provide a reliable reference for the final research results of this article.

(3) Mathematical Statistics

Use the relevant software to carry on the statistical analysis to the research result of this article.

4. Experimental Analysis of E-commerce New Media Marketing Based on Data Mining

4.1. Analysis of the Similarity between Clusters

In order to be able to understand the experimental results in a deeper level, we will carry out related quantitative and qualitative analysis on the experimental results of the Customer data set below. Currently, the selected parameters are K=5 and λ=0.5. NetSimilarity is an indicator for evaluating the quality of K-medodis clustering. It represents the sum of similarities between non-cluster center objects in the cluster and the cluster center. The larger the value, the better the clustering quality. The data obtained is shown in Table 1.

Table1. Similarity between clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	0.8574	0.1650	0.0824	0.0895	0.0937
Cluster 2	0.1650	0.8526	0.0969	0.1043	0.1224
Cluster 3	0.08235	0.0969	0.8647	0.0914	0.1135
Cluster 4	0.0895	0.1043	0.0914	0.8529	0.1391
Cluster 5	0.0937	0.1224	0.1135	0.08391	0.8391

It can be seen from Figure 1 that the element value on the diagonal represents the average of the similarity between each element in the cluster and the cluster center, while the element value on the non-diagonal line represents the similarity between the cluster centers of the two clusters. From the data shown in the table, the value of the elements on the diagonal is significantly greater than the elements on the off-diagonal line. This shows that the clustering results we obtained make the

elements in each cluster the most similar, and each cluster is the element with more obvious differences. By dividing the customers, we can define various types of consumers.

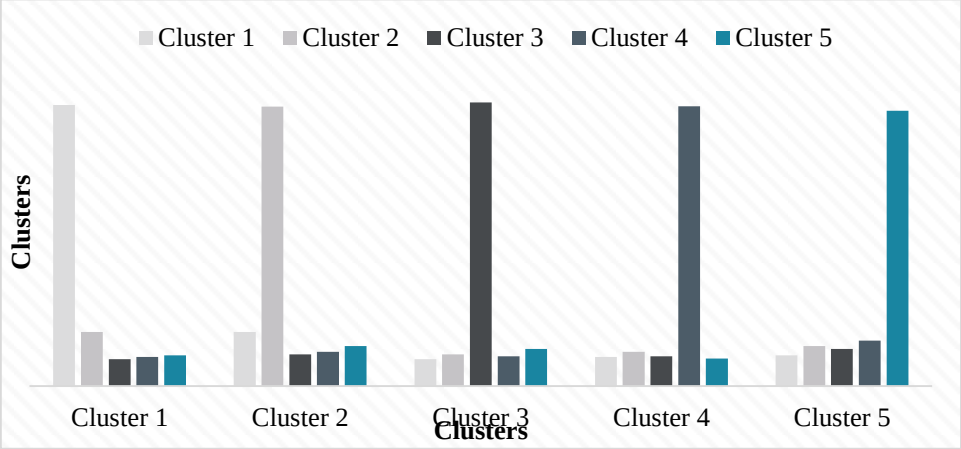

Figure 1. Similarity between clusters

4.2. Top-N Recommended Experiment Results and Analysis

The hit rate of each cluster center user during Top-N recommendation under each similarity method is recorded in turn, and finally the average hit rate corresponding to each method is calculated. The red bold value in the table represents the value with the highest recommendation rate in the current group of experiments. The results of the experiment are shown in Table 2.

Table 2. Top-N recommended experiment results and analysis

	Attribute SimRank	SimRank	Attribute
N=10	24.00%	14.00%	10.00%
N=20	15.00%	10.00%	12.00%
N=30	10.70%	9.30%	12.00%
N=40	9.00%	7.50%	9.00%
N=50	9.20%	7.20%	8.00%
N=60	8.00%	6.33%	7.00%
N=70	7.14%	5.71%	6.71%
N=80	6.25%	5.25%	6.25%
N=90	5.08%	4.60%	5.56%
N=100	5.80%	4.40%	5.20%

It can be seen from Figure 2 that when user similarity calculation adopts the method Attribute-SimRank in this paper, the average hit rate of product recommendation for cluster center users is in most cases higher than that of SimRank and Attribute methods. When the value of N changes from 10 to 100, only when N=30 and N=90, the Avg_HitRate (10.7% and 5.08%) corresponding to the Attribute-SimRank method is slightly lower than the Avg_HitRate (12% and 12% and 5.08%) corresponding to the Attribute method. 5.56%). In general, Attribute-SimRank increased the average hit rate of SimRank by about 2.6 percentage points, and Attribute-SimRank increased the average hit rate of Attribute by about 1.85 percentage points. Therefore, the results of this experiment further indirectly and effectively evaluate the similarity calculation method proposed in this paper. Because in the recommendation application, if the target user's similar interest user acquisition is more accurate, the recommendation result is more meaningful. The

accuracy of finding users with similar interest depends on whether the similarity acquisition between similar users is accurate.

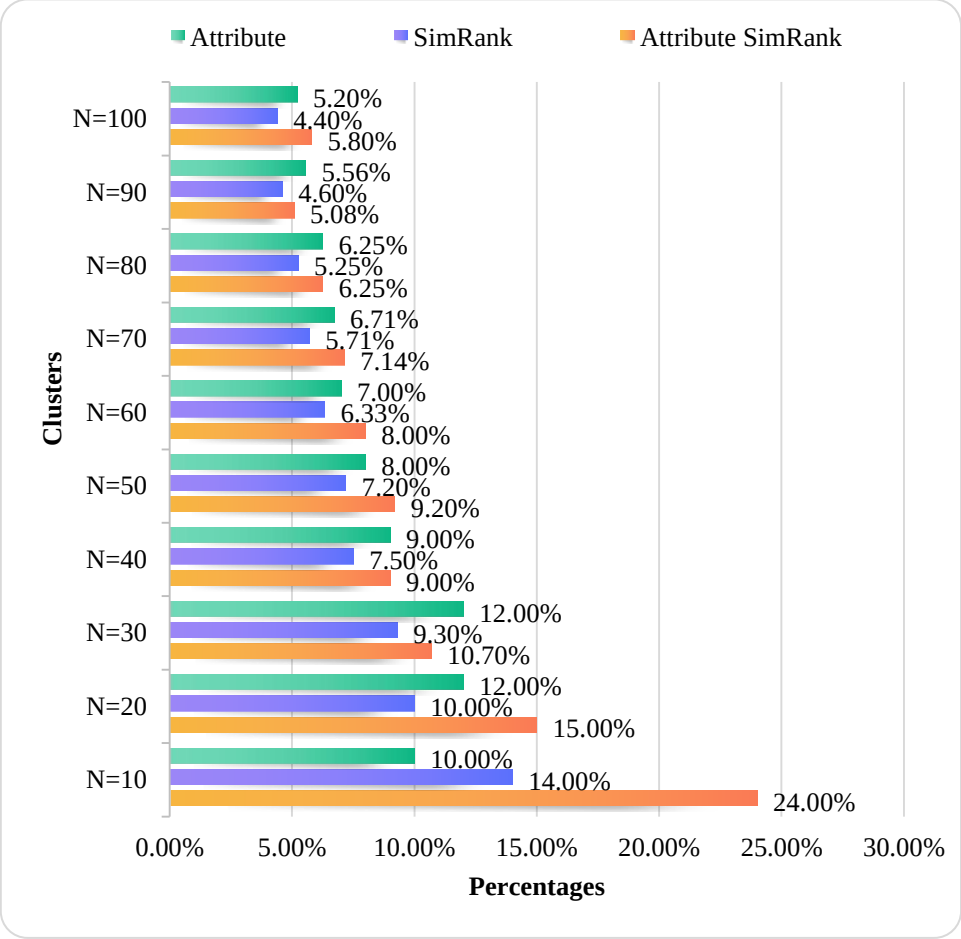

Figure 2. Top-N recommended experiment results and analysis

5. Conclusion

China's new media marketing has developed rapidly, and the number and scale of online shopping users have reached a new high. As a result, information related to customers, commodities, purchases, etc. has exploded, leading to the problem of information overload in the e-commerce field. Faced with the massive and complex e-commerce information, it is difficult for customers or market managers to quickly and accurately obtain the required information, which affects customer consumption and the decision-making process of managers. For users, whether they are new users or old users, faced with the increasingly abundant information on goods and services on the Internet, they will be confused about the choice of information; for businesses, when faced with so many users, how to master the user's preferences and buying habits Become the problem they most need to understand. To solve the problems of buyers and sellers in the process of online shopping, the data related to product sales can be fully utilized and in-depth mining.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Pappas I O, Kourouthanassis P E, Giannakos M N, et al. Explaining online Shopping Behavior With fsQCA: The role of Cognitive And Affective Perceptions. Journal of Business Research, 2016, 69(2):2016. https://doi.org/10.1016/j.jbusres.2015.07.010

[2] Yu W Y, Yan C G, Ding Z J, et al. Modeling and Verification of Online Shopping Business Processes by Considering Malicious Behavior Patterns. IEEE Transactions on Automation Science & Engineering, 2016, 13(2):647-662. https://doi.org/10.1109/TASE.2014.2362819

[3] Zhou L, Wang X, Lin N, et al. Location-Routing Problem with Simultaneous Home Delivery and Customer's Pickup for City Distribution of Online Shopping Purchases. Sustainability, 2016, 8(8):828-. https://doi.org/10.3390/su8080828

[4] Fang J, Wen C, George B, et al. Consumer Heterogeneity, Perceived Value, and Repurchase Decision-Making in Online Shopping: The Role of Gender, Age, and Shopping Motives. Journal of Electronic Commerce Research, 2016, 17(2):116-131.

[5] Agag G. Cultural and religiosity drivers and Satisfaction Outcomes of Consumer Perceived Deception In Online Shopping. Internet Research, 2016, 26(4):5549-5554. https://doi.org/10.1108/IntR-06-2015-0168

[6] Li Yanfei. Explore the "online Shopping era"% Explore the "online Shopping era" Commodity Packaging Design Innovation. Design, 2016, 000(7): 138-139.

[7] Shen Z, Hou D, Zhang P, et al. Lead-based paint in children's toys sold on China's major online shopping platforms.. Environmental Pollution, 2018, 241(10):311-318. https://doi.org/10.1016/j.envpol. 2018.05.078

[8] Lissitsa S, Kol O. Generation X vs. Generation Y - A decade of online shopping. Journal of Retailing and Consumer Services, 2016, 31(7):304-312. https://doi.org/10.1016/j.jretconser. 2016.04.015

[9] Chakraborty R, Lee J, Bagchi-Sen S, et al. Online Shopping Intention in the Context of Data Breach in Online Retail Stores: An examination of Older And Younger Adults. Decision Support Systems, 2016, 83(3):47-56. https://doi.org/10.1016/j.dss.2015.12.007

[10] Ding Y, Lu H. The Interactions between online Shopping and Personal Activity Travel Behavior: An Analysis with a GPS-based activity travel diary. Transportation, 2017, 44(2):1-14. https://doi.org/10.1007/s11116-015-9639-5

[11] Beuckels E, Hudders L. An Experimental Study to Investigate the Impact of Image Interactivity on the Perception of Luxury In An Online Shopping Context. Journal of Retailing & Consumer Services, 2016, 33(11):135-142. https://doi.org/10.1016/j.jretconser.2016.08.014

[12] Zhou Q, Xia R, Zhang C. Online Shopping Behavior Study Based on Multi-Granularity Opinion Mining: China vs. America. Cognitive Computation, 2016, 8(4):587-602. https://doi.org/ 10.1007/ s12559-016-9384-x